

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

1/5/2  
DIALOG(R)File 351:Derwent WPI  
(c) 2004 Thomson Derwent. All rts. reserv.

011767658 \*\*Image available\*\*

WPI Acc No: 1998-184568/ 199817

XRPX Acc No: N98-146412

Machine translation apparatus such as word processor - has sentence  
dividing part for segregating fed character data into predetermined  
sentences using additional information fed along input document

Patent Assignee: SHARP KK (SHAF )

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
-----------	------	------	-------------	------	------	------

JP 10040251	A	19980213	JP 96194450	A	19960724	199817 B
-------------	---	----------	-------------	---	----------	----------

Priority Applications (No Type Date): JP 96194450 A 19960724

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
-----------	------	-----	----	----------	--------------

JP 10040251	A		9	G06F-017/27	
-------------	---	--	---	-------------	--

Abstract (Basic): JP 10040251 A

The translation apparatus adds additional information indicating the character data, the start position and completion position of the input document (1) in the present language. This input information is fed to an input part. The fed input information is segregated into predetermined sentences by a sentence dividing part (2). Based on an additional information containing predetermined sentence paragraph symbol.

Then, the sequence of character data in the character string and the additional information is stored after which the character data and the additional information are separated by the tag isolation part (3). The isolated character data is then translated into the required language by a translation part (4). The tag is resorted to the translated data by a tag restoration part (5).

ADVANTAGE - Enables proper translation of original document.

Enables proper restoration of tags to translated information.

Dwg.1/15

Title Terms: MACHINE; TRANSLATION; APPARATUS; WORD; PROCESSOR; SENTENCE;  
DIVIDE; PART; SEGREGATE; FEED; CHARACTER; DATA; PREDETERMINED; SENTENCE;  
ADD; INFORMATION; FEED; INPUT; DOCUMENT

Index Terms/Additional Words: COMPUTER

Derwent Class: T01

International Patent Class (Main): G06F-017/27

File Segment: EPI

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-40251

(43) 公開日 平成10年(1998) 2月13日

(51) IntCl.<sup>6</sup>

G 0 6 F 17/27

識別記号

片内整理番号

F I

G 0 6 F 15/38

15/20

D

5 5 0 E

技術表示箇所

審査請求 未請求 請求項の数 3 O L (全 9 頁)

(21) 出願番号

特願平8-194450

(22) 出願日

平成8年(1996) 7月24日

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 西浦 一夫

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

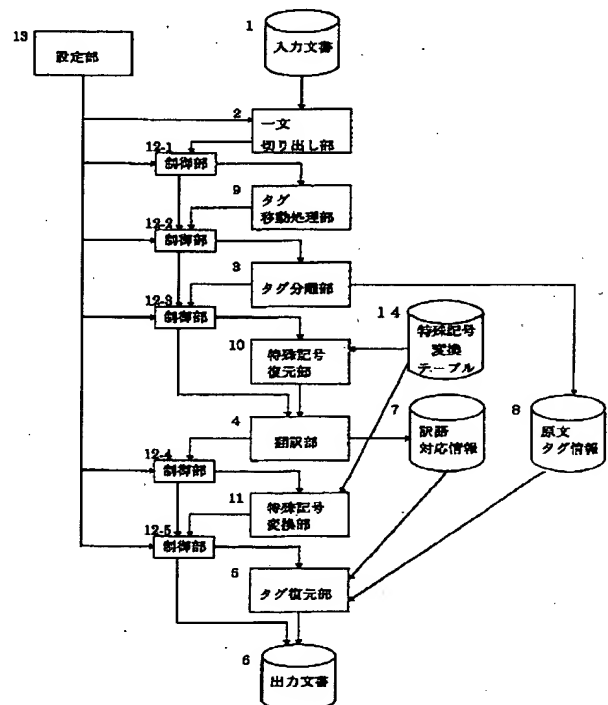
(74) 代理人 弁理士 梅田 勝

(54) 【発明の名称】 機械翻訳装置

(57) 【要約】

【課題】 タグ情報を含んだ文書を翻訳する際に、タグの中に文区切り記号がある場合、そこで文が区切られるとタグ情報が破壊されてしまい、翻訳が正常に行われな。さらに、単語中にタグが存在すると、その単語が分離されてしまい、単語の情報が正確に得られない。さらには、タグに使用される文字が、入力文及び本訳文に出現すると翻訳が正常に行われないという問題があった。

【解決手段】 タグの中にある文区切り記号をスキップする文区切り記号検出部と、単語の中にあるタグを移動するタグ移動処理部と、文中の特殊記号を復元及び変換する特殊記号復元部及び特殊記号変換部を備えることにより上記課題を解決する。



## 【特許請求の範囲】

【請求項1】 第1言語により記述された文字データおよび、開始位置及び終了位置を示す一対のタグにより、その区間において各種情報を付加する付加情報とからなる入力文を入力する入力部と、  
前記入力部より入力された文字データおよび付加情報を、所定の文区切り記号によって分割する一文切り出し部と、  
前記一文切り出し部によって切り出された文字列における、文字データと付加情報の関係を記憶し、前記付加情報と文字データを分離するタグ分離部と、  
前記タグ分離部によって分離された文字データを所望の第2言語に翻訳し、訳語の対応を記憶する翻訳部と、  
前記翻訳部によって翻訳された第2言語の訳文に、タグを復元するタグ復元部と、  
前記タグ復元部によって、復元された第2言語の訳文を出力する出力部を備えた機械翻訳装置において、  
前記一文切り出し部において、入力文中の所定の文区切り記号に付加情報が付加されている場合は、その区切り記号において分割を行わないことを特徴とする機械翻訳装置。

【請求項2】 第1言語により記述された文字データおよび、開始位置及び終了位置を示す一対のタグにより、その区間において各種情報を付加する付加情報とからなる入力文を入力する入力部と、  
前記入力部より入力された文字データおよび付加情報を、所定の文区切り記号によって分割する一文切り出し部と、  
前記一文切り出し部によって切り出された文字列における、文字データと付加情報の関係を記憶し、前記付加情報と文字データを分離するタグ分離部と、  
前記タグ分離部によって分離された文字データを所望の第2言語に翻訳し、訳語の対応を記憶する翻訳部と、  
前記翻訳部によって翻訳された第2言語の訳文に、タグを復元するタグ復元部と、  
前記タグ復元部によって、復元された第2言語の訳文を出力する出力部を備えた機械翻訳装置において、  
前記タグが一単語の途中に存在する場合において、そのタグを該単語の先頭あるいは末尾に移動させるタグ移動処理部を備えたことを特徴とする機械翻訳装置。

【請求項3】 第1言語により記述された文字データおよび、開始位置及び終了位置を示す一対のタグにより、その区間において各種情報を付加する付加情報とからなる入力文を入力する入力部と、  
前記入力部より入力された文字データおよび付加情報を、所定の文区切り記号によって分割する一文切り出し部と、  
前記一文切り出し部によって切り出された文字列における、文字データと付加情報の関係を記憶し、前記付加情報と文字データを分離するタグ分離部と、

前記タグ分離部によって分離された文字データを所望の第2言語に翻訳し、訳語の対応を記憶する翻訳部と、  
前記翻訳部によって翻訳された第2言語の訳文に、タグを復元するタグ復元部と、  
前記タグ復元部によって、復元された第2言語の訳文を出力する出力部を備えた機械翻訳装置において、  
前記翻訳部において、特定の意味を表わす特殊記号を、その本来の文字へと復元する特殊記号復元部と、  
前記特殊記号復元部によって復元された文字を、翻訳部による翻訳処理後に特殊記号へと変換する特殊記号変換部を備えたことを特徴とする機械翻訳装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】第1言語で記述され付加情報を含む文書を、所望の第2言語に翻訳する機械翻訳装置に関するものである。

## 【0002】

【従来の技術】機械翻訳装置を含め、ワードプロセッサやコンピュータ等の計算機上で取り扱われる文書データの多くは、文字データ本体である文字情報以外にも、文書の構成、レイアウトや文字の修飾情報、例えば野線の付与や書体の変更など各種の付加情報を含んでいる。

【0003】これら付加情報の方式としてはいくつかの規格が既に提案されており、文書の論理構造などを記述するためのSGML、ハイパーテキスト文書を記述するためのHTMLなどが知られている。

【0004】このような状況において、機械翻訳の分野においても、原文書のデータの中に含まれている付加情報をできるだけ損なわずに、通常の文書の部分を翻訳することが行われている。このような装置で翻訳された翻訳結果の文書は、原文書データとほとんど同様の書式で出力することが可能となっている。

【0005】このような機械翻訳装置において、開始タグ/終了タグと呼ばれる記号を用いて、特定の区間に付加情報を付与させている文の処理過程の例を次に示す。

【0006】Time flies *like an arrow* .

上記の文において、“*<italic>*”、“*</italic>*”がそれぞれ開始タグ、終了タグであり、これらの両タグはこの二つのタグでは含まれた区間をイタリック体で出力することを表している。

【0007】この文は以下に説明する処理手順によって、同じ付加情報を持った訳文へと変換される。図11の処理ブロック図を元に説明する。

【0008】まず、一文切り出し部2では、入力文書1中の文区切り記号（例えば“.”や改行記号等）を検出することにより、一文毎に文字列を次の処理へと受け渡す。

【0009】一文切り出し部2は、図12のブロック図に示されるような構成を持つ。操作部2-1では入力さ

れる文を先頭から順に走査し、2-2の文区切り記号検出部において所定の文区切り記号を検出する。検出された文区切り記号に基づいて、2-3の分割処理部において分割を行う。

【0010】タグ分離部3で、一文切り出し部1から受け取った文字列中に出現するタグを識別し、どの単語にタグ情報が付加されているのか記憶する。図13はタグ情報の記憶状態(原文タグ情報8)を示す図である。タグ情報を記憶後、タグ分離部3はタグを原文から分離する。

【0011】次に、翻訳部4によって、タグが削除された文の翻訳を行う。さらに、翻訳過程で得られる原文と訳文の単語の対応情報を記憶する。図14は訳語の対応情報の例(訳語対応情報7)を示す図である。

【0012】さらに、タグ復元部5は、翻訳部4で得られた翻訳結果、単語の対応情報とタグ分離部で記憶しておいた原文タグ情報をもとにして、新たに訳文にタグを付与する。図15が得られる訳文タグ情報を示す図である。これに基づいて訳文にタグを付加し、出力文書6が出力される。この例で得られる出力文書(翻訳結果)は次のようになる。この例では入力文における“like”、“an”、“arrow”に対応する訳語である“矢”、“のように”にタグが復元されて出力されている。

【0013】時間は<italic>矢のように</italic>飛ぶ。

【0014】

【発明が解決しようとする課題】しかしながら、タグ情報を含んだ文書の翻訳装置においては、上記のような処理を行っても必ずしも使用者の要求する出力結果が得られない場合がある。以下のその問題点を説明する。

【0015】まず、問題の一つは、複数の文より構成される原文書から、翻訳する文を一文ずつを切り出す処理に関する点である。現在の機械翻訳装置においては、翻訳を行う文書が複数の文から構成される場合には、文書から一文ずつ文を切り出す処理を行い、それぞれの文に対して順次翻訳処理を行うのが通常である。

【0016】これらの文の切り出しは、原文書中の句読点や改行など、装置によってあらかじめ定められた記号を区切りとして処理が行われる。また、文の区切りとする記号をユーザが任意に設定可能な装置も存在する。

【0017】ところがタグ情報を含む文書の翻訳装置においては、この切り出し処理が上記の方法ではうまく行えないことがある。例えば以下はHTMLで記述された文の例である。この例ではリンク先のアドレスをタグ内に記載している。

【0018】Click <a href = http://www.sharp.co.jp> here </a> !

このような例において、例えば文の切り出し記号として“:”が設定されていると、

Click <a href = http:

の部分で文の切り出しが行われてしまう。よって、この文の翻訳が正常に行えなくなってしまう。また、翻訳が正常に行われただけでなく、タグ情報も分離により破壊されてしまう。

【0019】別の問題点としては、以下のような例である。

【0020】C<font size=1>lick here</font> !

これは文の先頭以外の文字のフォントサイズを異ならせ、文書表示の際のレイアウトを工夫したHTMLタグの記述の一例である。タグによりClickの単語の先頭文字以外のフォントを変更している。

【0021】この文の翻訳を、前記した従来技術によって処理することを考える。従来の技術においては、処理の手順から明らかなようにタグ情報の扱いは単語単位が前提となるから、この文の処理においては本来1単語である“Click”を“C”と“lick”の2単語とし、単語“lick”以降にタグが付与されているとして、扱われることとなる。そのためタグ情報は残されるものの、翻訳結果として正常な訳文は得られなくなってしまう。

【0022】その他、タグ情報を含む文書の翻訳の際の問題として、タグの仕様に基づく特殊記号の扱いがある。例えば、HTMLの仕様ではタグの開始記号、終了記号として“<”、“>”が定義されている。そのため、文書の作成者が文中に“<”や“>”の記号を利用したい時には、これをタグの開始、終了記号と区別するために、“&lt;”、“&gt;”を代わりに用いることになっている。そのため、

<font size=10>C</font>lick &lt; here! &gt;

といった文を翻訳する際には、単純にタグを除去して翻訳するだけでは望ましい翻訳結果は得られない。

【0023】また、原文の翻訳結果として訳文中に“<”や“>”の記号が含まれる場合には、翻訳した文書をブラウザ等の表示処理により表示を行う際にこれらの記号がタグ記号と判断されてしまい、表示が翻訳結果と異なってしまうという問題もある。

【0024】従来においてタグ情報を含む文書の翻訳の際にはこれらが原因となって、タグを含まない文書に比べて翻訳の精度が低下してしまっていた。また逆に、タグ処理を行う翻訳装置でタグを含まない文書の翻訳をおこなった際には、タグ処理の副作用のために本来の翻訳結果と異なってしまうこともあった。

【0025】

【課題を解決するための手段】本発明の請求項1によれば、第1言語により記述された文字データおよび、開始位置及び終了位置を示す一対のタグにより、その区間において各種情報を付加する付加情報とからなる入力文を入力する入力部と、前記入力部より入力された文字データおよび付加情報を、所定の文区切り記号によって分割する一文切り出し部と、前記一文切り出し部によって切

## 5

り出された文字列における、文字データと付加情報の関係を記憶し、前記付加情報と文字データを分離するタグ分離部と、前記タグ分離部によって分離された文字データを所望の第2言語に翻訳し、訳語の対応を記憶する翻訳部と、前記翻訳部によって翻訳された第2言語の訳文に、タグを復元するタグ復元部と、前記タグ復元部によって、復元された第2言語の訳文を出力する出力部を備えた機械翻訳装置において、前記一文切り出し部において、入力文中の所定の文区切り記号に付加情報が付加されている場合は、その区切り記号において分割を行わないことにより上記課題を解決する。

【0026】本発明の請求項2によれば、第1言語により記述された文字データおよび、開始位置及び終了位置を示す一対のタグにより、その区間において各種情報を付加する付加情報とからなる入力文を入力する入力部と、前記入力部より入力された文字データおよび付加情報を、所定の文区切り記号によって分割する一文切り出し部と、前記一文切り出し部によって切り出された文字列における、文字データと付加情報の関係を記憶し、前記付加情報と文字データを分離するタグ分離部と、前記タグ分離部によって分離された文字データを所望の第2言語に翻訳し、訳語の対応を記憶する翻訳部と、前記翻訳部によって翻訳された第2言語の訳文に、タグを復元するタグ復元部と、前記タグ復元部によって、復元された第2言語の訳文を出力する出力部を備えた機械翻訳装置において、前記タグが一単語の途中に存在する場合において、そのタグを該単語の先頭あるいは末尾に移動させるタグ移動処理部を備えたことにより上記課題を解決する。

【0027】本発明の請求項3によれば、第1言語により記述された文字データおよび、開始位置及び終了位置を示す一対のタグにより、その区間において各種情報を付加する付加情報とからなる入力文を入力する入力部と、前記入力部より入力された文字データおよび付加情報を、所定の文区切り記号によって分割する一文切り出し部と、前記一文切り出し部によって切り出された文字列における、文字データと付加情報の関係を記憶し、前記付加情報と文字データを分離するタグ分離部と、前記タグ分離部によって分離された文字データを所望の第2言語に翻訳し、訳語の対応を記憶する翻訳部と、前記翻訳部によって翻訳された第2言語の訳文に、タグを復元するタグ復元部と、前記タグ復元部によって、復元された第2言語の訳文を出力する出力部を備えた機械翻訳装置において、前記翻訳部において、特定の意味を表わす特殊記号を、その本来の文字へと復元する特殊記号復元部と、前記特殊記号復元部によって復元された文字を、翻訳部による翻訳処理後に特殊記号へと変換する特殊記号変換部を備えたことにより上記課題を解決する。

【0028】

【発明の実施の形態】本発明の実施の形態を図面を用い

## 6

て詳細に説明する。図1は本発明の構成を示すブロック図である。従来技術として示した図11に比べてタグ移動処理部9、特殊記号復元部10、特殊記号変換部11、および制御部12-1から12-5、設定部13、特殊記号変換テーブル14が追加されている。

【0029】また、一文切り出し部2の処理が従来の一文切り出し部とは異なっている。図2は本発明における一文切り出し部2の詳細な構成を示すブロック図であるが、従来の一文切り出し部のブロック図12と比べてタグ記号処理部2-4、制御部2-5が追加されている。入力文書1、タグ分離部3、翻訳部4、タグ復元部5、出力文書6、訳語対応情報7、原文タグ情報8は図11で示した従来技術のものと同様の構成である。

【0030】一文切り出し部2では、入力文書1中の文区切り記号を検出することにより、一文毎に文字列を次の処理へと受け渡す。一文切り出し部2の処理の流れを示す図が図3である。一文切り出し部2は、メモリーやファイルに存在する入力文書から、定められた文区切り記号を検出することにより、入力文書を一文単位で次の処理に受け渡す処理を行う部分である。

【0031】図3において、ステップS1、ステップS5、ステップS6、ステップS7、ステップS8は一文切り出し部で従来より行われる処理である。走査位置の文字を検査して(ステップS1)、文区切り記号であれば(ステップS5)、現在の走査位置までを一文の区切りとして次の処理部に受け渡す(ステップS6)。走査位置を次々と進め(ステップS7)、以上の処理を文書の終わりまで繰り返す(ステップS8)。

【0032】ステップS2、ステップS3、ステップS4が本発明でタグ記号処理部2-4の追加により行われる処理である。現在の走査位置がタグ開始記号であれば(ステップS2)、タグ終了記号が出現するまで(ステップS4)、走査をスキップする(ステップS3)。

【0033】この処理の追加によりタグ中の文区切り記号が検出されることがなくなり、タグの途中で文が分割されることなく次の処理へと文字列を受け渡すことが可能となる。

【0034】図4は入力文字列の例である。先頭から順に走査を行い、タグの開始記号を検出すると、次のタグの終了記号を検出するまで、区切り記号の処理は行わない。つまり、図4の例では“<”が検出されたあと、“>”が検出されるまで、区切り記号“:”や“.”が出現しても文の区切りを行わない。

【0035】次に図1のタグ移動処理部9の流れを図5に示す。以下、図5のフローチャートに従って説明を行う。タグ移動処理部9においては、まず一文切り出し部2から受け渡された文字列を図示しない一時的なバッファにセットする(ステップS11)。次に、セットしたバッファを先頭から走査し、タグの検出を行う(ステップS12)。タグの検出が行われたら、そのタグが一単

語中に割り込んで出現しているかを判断する。検出したタグの直前の文字及び直後の文字を調べ、これらが共に単語区切り記号でない場合には、タグが一単語中に割り込んでいると判定する(ステップS13、S14)。単語区切り記号とは空白記号や改行記号に代表されるような、単語と単語の区切りとして使用される記号である。

【0036】タグが一単語中に割り込んでいると判定された時には、そのタグが開始タグか終了タグかを調べ(ステップS15)、開始タグならばタグの前方へバッファの走査を行い、単語区切り記号より単語の先頭位置を検出して(ステップS16)、その位置へとタグを移動させる(ステップS18)。タグが終了タグならばタグの後方へ走査を行い、単語区切り記号より単語の末尾を検出して(ステップS17)、その位置へとタグを移動させる(ステップS18)。

【0037】図6および図7は、タグ移動処理部での処理実行例である。図6は図5におけるステップS11で文字列のセットを行った直後のバッファの状態を示している。この文字列に対して、ステップS12でバッファ中のタグ<b>および</b>が検出される。

【0038】次にそれぞれのタグについてステップS13、S14でタグの前後の文字を検査するが、タグ<b>については、タグの直前が単語区切り記号(空白記号)なのでタグの移動処理は行われない。タグ</b>は前後の文字共に単語区切り記号でないのでタグの移動処理が行われる。</b>は終了タグであるため、タグ後方の単語区切り記号を調べ(ステップS17)、その位置までタグを移動する(ステップS18)。タグの移動を行った後のバッファの状態が図7である。

【0039】図6では“here”の単語の中にタグ“</b>”が割り込んで“<b>h</b>ere”となっている。割り込んでいるタグ“</b>”は終了タグなので、次の単語区切り位置である“here”と“!”の間にタグ“</b>”を移動させる。

【0040】タグ分離部3では、一文切り出し部1から受け取った文字列中に出現するタグを識別し、どの単語にタグ情報が付加されているのか記憶する。タグ情報を原文タグ情報8として記憶後、タグ分離部3はタグを原文から分離する。

【0041】次に、図1の特殊記号復元部10、特殊記号変換部11、特殊記号変換テーブル14について説明する。特殊記号復元部10の処理の流れを図8に、特殊記号変換部11の処理の流れを図9に示す。

【0042】特殊記号復元部10では、タグ分離部3からタグ記号が完全に分離された文字列を受け取り、この文字列中に出現する特殊記号を本来の記号へと復元する処理を行う。

【0043】特殊記号の復元処理は、特殊記号変換テーブル14を参照して行う。図10は特殊記号変換テーブル14の一例であり、特殊記号と、本来の記号とが対応

付けてある。

【0044】図8のステップS21で、受け取った文字列を図示しない一時バッファにセットし、ステップS22で文字列照合により特殊記号の検出を行い、特殊記号変換テーブルに基づいて特殊記号の復元を行う(ステップS23)。復元後の文字列はタグや特殊記号を一切含まない通常の文として次の翻訳部4へと受け渡される。

【0045】例えば、“<b>LAN</b> &gt; local area network &lt;”という入力文においては、まずタグ分離部3により、“<b>”と“</b>”のタグがまず分離され、特殊記号復元部10に文字列が引き渡される。特殊記号復元部10では所定の特殊記号が特殊記号変換テーブルによって変換され、ここでの最終的な出力は“LAN<br>local area network >”となる。

【0046】次に、翻訳部4によって、タグが削除された文の翻訳を行う。さらに、翻訳過程で得られる原文と訳文の単語の対応情報を訳語対応情報7として記憶する。

【0047】特殊記号変換部11では、図9に示すように、翻訳部4から翻訳結果を受け取り、バッファにセットし(ステップS31)、この文字列中に出現するタグ記号として使用される記号等を検出し(ステップS32)、特殊記号へと変換する(ステップS33)処理を行う。この処理は特殊記号復元部10と同様に特殊記号変換テーブル14に基づいて行われる。

【0048】例えば、翻訳部4から“セットする<ステップ1>”という訳文が出力された場合、“<”“>”の記号がそれぞれ“&st;”“&lt;”に変換され、“セットする&st;ステップ1 &lt;”と出力される。

【0049】変換後の文字列は次のタグ復元部5へと受け渡され、タグ復元部でタグ情報が復元される。その結果出力される出力文書6はタグの仕様に基づくタグ記号、及び特殊記号に従った翻訳結果となる。

【0050】さらに、以上で説明したタグに関する処理を行う各部に対してそれぞれ制御部12-1~12-5、2-5を設け、設定に応じて、タグに関する処理を行う/処理を行わずに次の処理部へ受け渡す、という処理の振り分けを行うことでタグを含む文書、タグを含まない文書のどちらについても適切な翻訳処理を行うことが可能となる。

【0051】図1の制御部12-1から12-5、および図2の制御部2-5が上記処理を行う部分であり、設定部13での設定内容を受けて処理の振り分けを行う。設定部13は、ユーザがインターフェイスを通じて明示的に設定するものでも良いし、入力文書を元に装置が自動的に設定を行うようなものでもよい。

【0052】

【発明の効果】本発明によれば、タグを復元することのできる機械翻訳装置において、タグ記号中の不本意な箇所で文が分割されて、翻訳が正常に行われなくなる事を

防ぐことができる。さらに、文中の一単語がタグにより分割され、二つの単語として認識されてしまうことを防止することが可能となる。

【0053】また、タグの仕様に基づく特殊記号が文書中に出現する際に、特殊記号をそのまま翻訳してしまい、翻訳の精度低下の原因となることを防ぐ。また、翻訳結果中にタグ記号と重複する記号が出現する場合において、これらを特殊記号に変換してから訳文の出力を行うことで、翻訳結果の表示の際にこれらの記号がタグとして扱われてしまうという問題を回避できる。

【図面の簡単な説明】

【図1】本発明の一実施の形態における構成を示すブロック図である。

【図2】本発明の一実施の形態における一文切り出し部2の詳細な構成を示すブロック図である。

【図3】本発明の一実施の形態における処理の流れを示すフローチャートである。

【図4】一文切り出し部での入力文書の走査状態例を示す図である。

【図5】本発明の一実施の形態における処理の流れを示すフローチャートである。

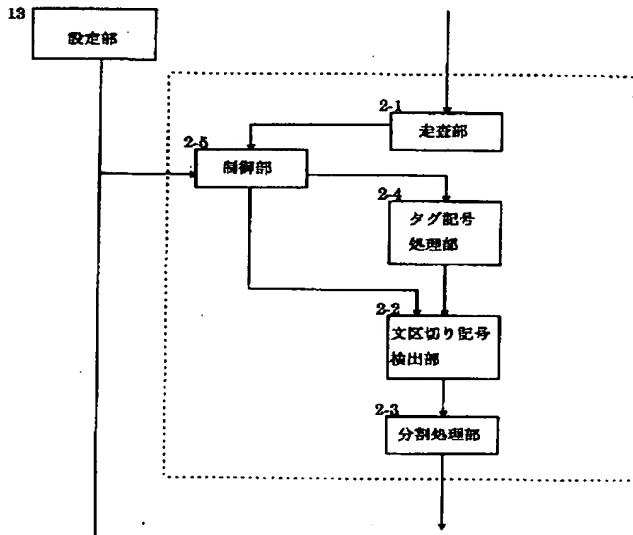
【図6】タグ移動処理部で処理を行う前の文字列の状態を示す図である。

【図7】タグ移動処理部で処理を行った後の文字列の状態を示す図である。

【図8】本発明の一実施の形態における処理の流れを示すフローチャートである。

【図9】本発明の一実施の形態における処理の流れを示

【図2】



すフローチャートである。

【図10】特殊記号復元部、特殊記号変換部で用いる特殊記号変換テーブルの一例である。

【図11】従来技術の構成を示すブロック図である。

【図12】従来技術における一文切り出し部2の詳細な構成を示すブロック図である。

【図13】従来技術におけるタグ情報の記憶例を示す図である。

【図14】従来技術におけるタグ情報の記憶例を示す図である。

【図15】従来技術における訳文タグ情報の記憶例を示す図である。

【符号の説明】

- 1 入力文書
- 2 一文切り出し部
- 3 タグ分離部
- 4 翻訳部
- 5 タグ復元部
- 6 出力文書
- 7 訳語対応情報
- 8 原文タグ情報
- 9 タグ移動処理部
- 10 特殊記号復元部
- 11 特殊記号変換部
- 12 制御部
- 13 設定部
- 14 特殊記号変換テーブル

【図6】

C l i c k < b > h < / b > o r e !

【図7】

C l i c k < b > h o r e < / b > !

【図10】

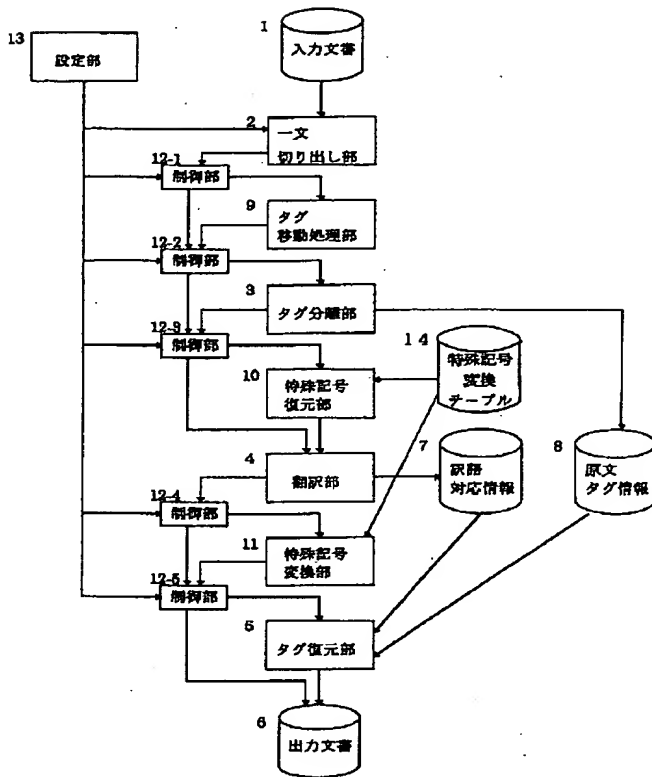
特殊記号	本来の記号
&lt;	<
&gt;	>
.....	.....

【図13】

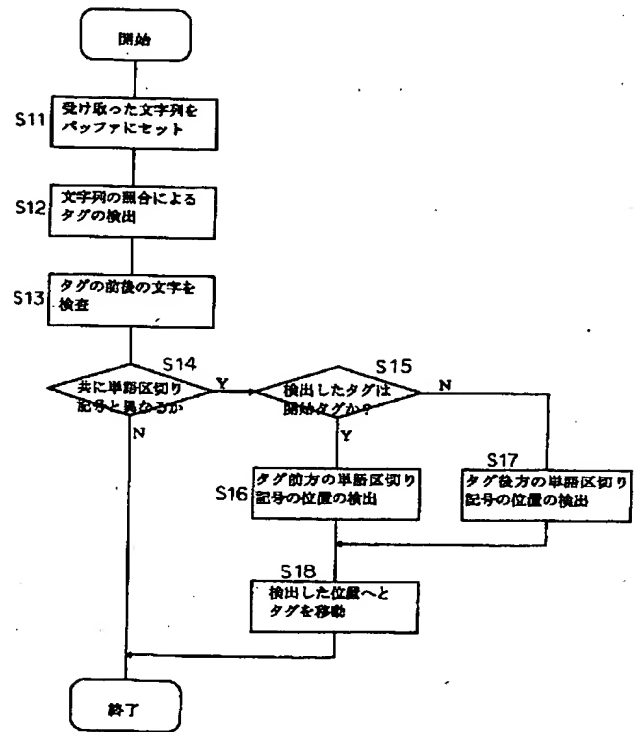
Time	
flies	
like	italic
an	italic
arrow	italic



【図1】



【図5】

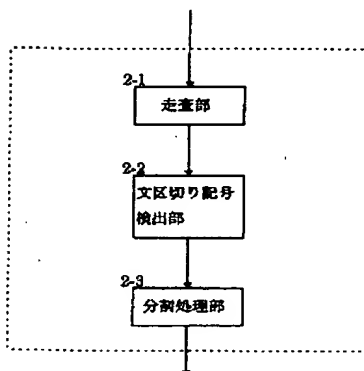


【図4】

Click <a href = http://www.sharp.co.jp> here </a> !

← この間では区切り記号の検出は行われずスキップされる → | ↔ |

【図12】



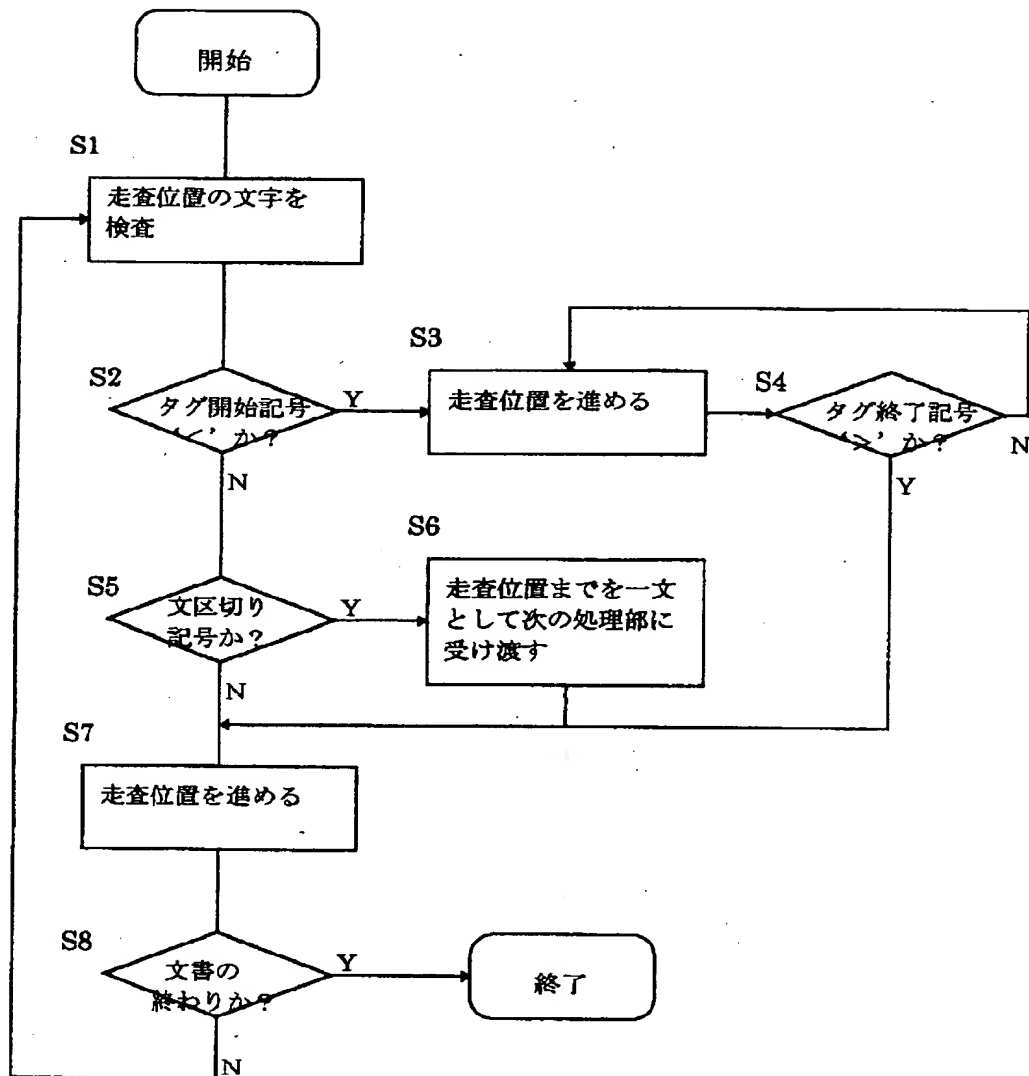
【図14】

Time	時間
flies	飛ぶ
like	のように
an	
arrow	矢
.	。

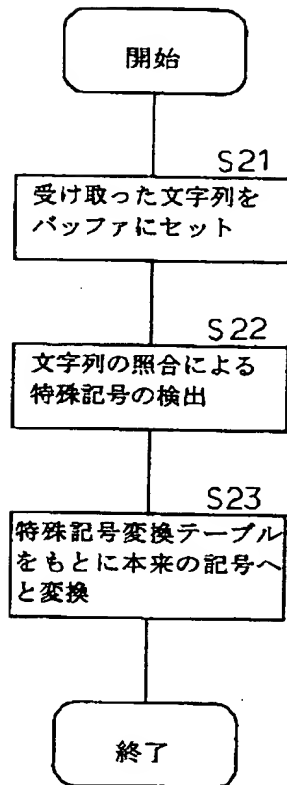
【図15】

時間	
矢	italic
のように	italic
飛ぶ	
。	

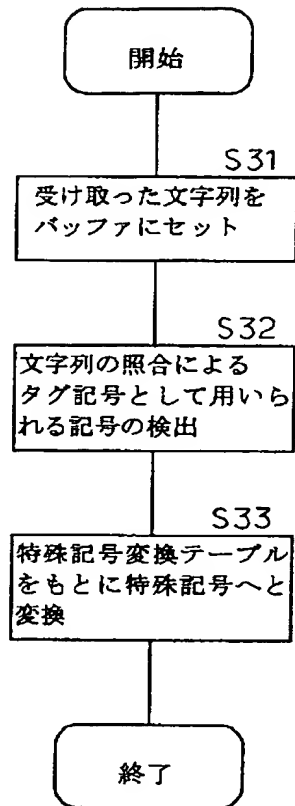
【図3】



【図8】



【図9】



【図11】

